

1/19

A)

	Motif 1	Motif 2	Motif 3	Motif 4
Rv2837c_Mc	VGWVCHVHPTDITICAGLALALVLDGG	VDLVTVVTHPSVDRLGALG	RELVTVDHPSND	SADSTTTVAETILDANGKPTDPRVAHCTIYAGLADTGSFRAASV
sl11253_Ssp	DLILCHOTACPDVGLGAAGLAKHPSGR	IRSELYVNOGDRGKAA	ROVATVDHNSP	AVGASTTLIVEKLQRADISLMSWEASVMAIGHVDTCGLTFTQT
MGPA_Mg	IVTFHVRPDCCLGAQOGLFLIKANF	EALATVDPANYNRIELRE	KAVLRDHPNED	SYVACEQIVENATVAKWTIPPVATLLYIGIYTCNRFYSNT
YTOI_Bs	IIILHRVRPDPAYGSCQGLTEILRETY	GALVTVCDTANQERIDDDQR	AKLMKDHPNED	SVSEMIYELYLEGKEHGMKLTQKAEELIYAGIYVGTGRFLPNT
AF2029_Af	LGFTFHDNPDPSMSSAYALREIAKQFD	YDAFAIVSSGPGVNNISIP	DISIVDHPFAEK	DVGATATILTEYIKELKITPSKILATLFFGIKSEDTDFEKNTR
MJ0977_Mj	NKILIVTHIDTPGLTSRAILQKLABRLN	YDLTFADLSSGQLMIKE	DKILIIDHPPEE	GAEICGAGVSYLPKALINNDWIDLAKYAVLGAVGCIQNIIEKLI
MJ1198_Mj	RPPIIRHRRDTPGYCGGIALEKALPII	LPLTVLDSNMLNLAERYL	IEVIVDHPFGE	KGRYDYREYLEKIALCWDFAFLRFMDGKGVDDIILATNIKEP
HP1042_Hp	MOYTHLSHTELGGYACQVSKQFPNIQ	EFLVDSNMLNLAERYL	IQIQLDHPISGK	IVYEFLKKHYAILEPKNTTLEPLVEMVNSVDIWTQCYGFELG
RecJ_Hi	QKIVIVGDPAAGATSTALSVLALRQLG	VQLLMTVNGVSSDFGVAF	IRVLVDHPPE	LAVRAKPFRELGIPTAETQPNFTDLDLVALGTLGAVPLDQNNR
RecJ_Hp	TEILVWGDYDADGVITSSAIMAKFPESLN	APLITVDNGINAFEAARF	YTLITDHPCLHH	LVAFYLCYGIHQLLGKESHSSELCLAGVATIAKMPPLTTFNR
RecJ_Ssp	EKVTIWGDPAAGITSTAVLWEGIGQFF	TKLIVTDTGSTNLDEIVY	MDVIVDHPFTUPD	VAFKLVEALYNQYPTVPQOPLDLDLVAIGLIADLVTLQGDOR
YYBQ_Bs	ILIEGHQNPETITICSAIYADLANKLG	VNGVILVCHNERQQSIKDI	QVLEVIDHERIAN	PVGCTATTILANKMYKENNVKIERKEIAGLMLSAIISDLSLLFKSPTC
ICRA_Sg	ILVFGHQNPSPAIGSSYAFAYLAREAY	AEQVILTCHNEFQQSVADI	EVYGVDDHERVAN	PVGSASSIVYRMFKHSVAVSKEIAGLMLSGUISDTLLKSPPT
Y606_Mj	RYVVGHNEDTPSIAIAIVLAYFLDCYP	GKEILVCHSEKSSQSFDDL	KLTATDHPKVGL	IAELYFKDAIDLIGCKKKELKPDLAGLILLISATISDTVLFKSPPT
AF0756_Af	VYVVGHNPDTPSVCSAIAFAYLWKNWK	GKNVALVCHSEKQATVDGI	EVVALVDPHEKIGD	PVGCTATVIKLLFPDKTQVEIPKDIAGILLISATISDTVLFKSATI
U60409_Lm	TVVQNGEGDTPSIVGCIYAMLFDKQP	QIAHNVDVIAALNASVLY	RVVGVDDHPDEQ	LRTVGSACTLVTELYRECGEDVVCPTLLTAPIVDTVNFPEPAQK
PPX1_SC	TICVGNESADTPSIAIAITYSCQYIYN	EINSLVYNNDTPKNLKNY	NVVGEDHPFDLQ	SCSSIVFNFWYKELQGDREVVMNIAPLILAGAILIDTSMNRKVE
PRUNE_Dm	HLVWGNESCELDPAVSNVTLAFVYAQRH	DVNVLVCHRVSPAPNVT	NVTEILDHPLED	SVGSCATLVAQRYLAEDQPRSTSVAQILLIATIVLDTINPAPAAK
h-prune.1_Hs	HVLGNBACCLDSTVSALALAFYLAKTT	QULTILVCHILSKSDTAL	AVAEDVDHPPIEP	LVGSCATLVTERILQGAPEILDORTAALLHGTITIDTCVNDMLKIG

h-prune.1_Hs

Specific motifs

Gis

Rv2837c_Mc	[107]	TVNLAVASFGCGGHRLLAGYTTGS	1646883
sl11253_Ssp	[123]	DTDLTQLLEPYGGGHAQAANLDRV	1653244
MGPA_Mg	[102]	GINVRDIAIKYGGGHNASCAITNK	1045875
YTOI_Bs	[103]	GPVINGLARKYNGGHPLASGASTYSW	2293259
AF2029_Af	[102]	EVLRRAFGDVGSAGGHAHAGAQTPLG	2648507
MJ0977_Mj	[254]	AIKYASEKVGSGGCHKFACGAYIPDN	2128614
MJ1198_Mj	[109]	QIMBEIPEASLDGGEHCAGSLKPEEG	2128720
HP1042_Hp	[137]	CDVCELSQWCFNGGHRNASGGKIDGP	2314198
RecJ_Hi	[168]	RIKSHQPNMILKFGGHMAAGLSIRRE	1172895
RecJ_Hp	[155]	DALNGVSSLLGCGGHRQAAGLSVEKN	2313437
RecJ_Ssp	[165]	ALLHSQRHMLGFGGHPFAAGLSPLD	1652638
YYBQ_Bs	[31]	DLSKKTVEELISLDAKEFTLG	586817
ICRA_Sg	[31]	NLASKSREELIDIDAKTPELN	1743856
MJ0608_Mj	[32]	VWGKLKPEEITNNDKFNPDEN	1591318
AF0756_Af	[32]	AVDDLTAHMDITKRDYKDFDMS	11498362
U60409_Lm	[36]	DVLALSVPQILRDYKQPSFK	1407725
PPX1_SC	[41]	DIKGFVSVDILKNDYKQPNFQ	730369
PRUNE_Dm	[37]	DISKLTLEVLKDMKVLQTD	1079081
h-prune.1_Hs	[34]	DVSLTTEQNLKRDQKTIYRQ	11245938

Family 2

Family 1

Fig. 1

04308203